University of Nebraska - Lincoln
DigitalCommons@University of Nebraska - Lincoln

Papers in Plant Pathology | Plant Pathology Department

1-1-2002

Mapping of the P1 proteinase cleavage site in the polyprotein of *Wheat streak mosaic virus* (genus *Tritimovirus*)

Il-Ryong Choi
University of Nebraska - Lincoln

Kempton M. Horken
University of Nebraska - Lincoln, khorken2@unl.edu

Drake C. Stenger
University of Nebraska - Lincoln

Roy C. French
University of Nebraska - Lincoln, rfrench2@unl.edu

Follow this and additional works at: http://digitalcommons.unl.edu/plantpathpapers

Part of the Plant Pathology Commons

Choi, Il-Ryong ; Horken, Kempton M.; Stenger, Drake C.; and French, Roy C., "Mapping of the P1 proteinase cleavage site in the polyprotein of *Wheat streak mosaic virus* (genus *Tritimovirus*)" (2002). *Papers in Plant Pathology.* Paper 11.
http://digitalcommons.unl.edu/plantpathpapers/11

This Article is brought to you for free and open access by the Plant Pathology Department at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in Papers in Plant Pathology by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.

Journal of General Virology (2002), **83**, 443–450. Printed in Great Britain

Mapping of the P1 proteinase cleavage site in the polyprotein of *Wheat streak mosaic virus* (genus *Tritimovirus*)

Il-Ryong Choi, Kempton M. Horken, Drake C. Stenger and Roy French

United States Department of Agriculture – Agricultural Research Service and Department of Plant Pathology, University of Nebraska, Lincoln, NE 68583, USA

Monopartite members of the family *Potyviridae* utilize three virus-encoded proteinases to cleave the viral polyprotein into mature proteins. The amino-terminal region of the viral polyprotein is autolytically cleaved by the P1 proteinase. A domain required for P1 proteinase activity of *Wheat streak mosaic virus* (WSMV) was mapped using a series of templates with nested 3′-truncations or 5′-deletions to program *in vitro* transcription–translation reactions. The WSMV P1 proteinase cleavage site was mapped to a position downstream of amino acid residue 348 and upstream of amino acid residue 353, with the peptide bond between amino acid residues Y_{352} and G_{353} the most probable site of hydrolysis. An alignment of potyvirus polyprotein sequences in the carboxy-terminal region of the P1 domain revealed WSMV P1 contained conserved H_{257}, D_{267}, S_{303} and $FIVXG_{325–329}$ residues upstream of the cleavage site that are typical of serine proteinases and shown by others to be required for P1 proteolysis in *Tobacco etch virus*. Insertion of the GUS reporter gene immediately downstream of the P1 cleavage site in a full-length clone of WSMV resulted in systemic infection and GUS expression upon inoculation of plants with *in vitro* transcripts. When cleaved by P1 at the amino terminus and NIa proteinase at a site engineered in the carboxy-terminus, active GUS protein expressed by WSMV in infected wheat had electrophoretic mobility similar to wild-type GUS protein.

Introduction

Wheat streak mosaic virus (WSMV) is the type species of the genus *Tritimovirus* within the family *Potyviridae*. WSMV has a genome organization typical of monopartite members of the family with a 9384 nucleotide (nt) RNA genome encoding a single large open reading frame (Stenger *et al.*, 1998). As with other potyviruses, the WSMV polyprotein is cleaved into mature proteins by viral-encoded proteinases. Evidence of proteinase activity by the NIa and HC-Pro proteins of WSMV has been presented (Choi *et al.*, 2000a, b), indicating that these proteins contribute to polyprotein processing in a fashion similar to that demonstrated (Carrington & Dougherty, 1987; Carrington *et al.*, 1989) for the most thoroughly investigated potyvirus, *Tobacco etch virus* (TEV).

TEV encodes a serine proteinase (P1) that acts in *cis* to cleave itself from the amino-terminal domain of the polyprotein (Verchot *et al.*, 1991, 1992). TEV P1 proteinase cleavage occurs between polyprotein residues Y_{304} and S_{305} (Verchot *et al.*, 1991). Because the WSMV polyprotein contains YS at positions 283 and 284, Stenger *et al.* (1998) speculated that WSMV P1 proteinase cleavage may occur between these two residues. Insertion of 12 non-viral nucleotides immediately downstream of this putative cleavage site into an infectious clone of WSMV had no effect on viability but insertion of a bacterial glucuronidase (GUS) gene at the same site abolished infectivity (Choi *et al.*, 2000b). One explanation for the aforesaid observations is that WSMV P1 proteinase cleavage occurs at a site other than Y_{283}–S_{284}, making it of interest to define the cleavage site experimentally. In this report, we describe coupled *in vitro* transcription–translation experiments designed to map the WSMV polyprotein domain involved in P1 proteolysis. We further demonstrate that the WSMV P1 proteinase cleavage site is processed *in vivo* such that insertion of GUS immediately downstream of the P1 proteinase cleavage site in an infectious clone of WSMV results in the accumulation of excised GUS protein in tissues of infected plants.

Author for correspondence: Roy French.
Fax +1 402 472 4020. e-mail rfrench@unlnotes.unl.edu

Methods

■ ***In vitro* transcription–translation.** DNA fragments corresponding to WSMV-Sidney 81 nt positions 131 to 985 (pCP1-F) and 639 to 2340 (pCSP1-8) were previously cloned in pACT2 (Choi *et al.*, 2000a).

0001-8059

Table 1. Oligonucleotide primers used in this study

Name	Nucleotide sequence (5′–3′)*	Cloning site (bold)	Underlined sequences identical to:
X-NPT5	GAGG**CTCGAG**GGGTTGAACAATGGA	*Xho*I	NPT II 4–18
NPT-3	GAGG**GTCGAC**GAAGAACTCGTCAAGAAG	*Sal*I	NPT II complementary strand 775–792
P1-5	CC**GGATCC**CAATGGCAACAGCGAATTGT	*Bam*HI	WSMV cDNA 131–148
P1-X21	CCTT**CTCGAG**TTTCGGCTTCAGTGCATT	*Xho*I	WSMV cDNA complementary strand 1145–1162
P1-X22	GGCG**CTCGAG**TCCATGTGTGCATTTCGG	*Xho*I	WSMV cDNA complementary strand 1157–1174
P1-X23	GAAG**CTCGAG**ATACCATCTAAGTCCATG	*Xho*I	WSMV cDNA complementary strand 1169–1186
P1-X24	CACC**CTCGAG**TGCTTGGTCGCCATACCA	*Xho*I	WSMV cDNA complementary strand 1181–1198
P1-X25	GTTG**CTCGAG**AACTTTGTTCACTGCTTG	*Xho*I	WSMV cDNA complementary strand 1193–1210
T7-ACT	*TAATACGACTCACTATAGGGAGACCACATG*GATGATGTATATAACTATCTATTC		pACT2 5117–5140
P1-3	CC**GGATCC**CTAATAAGTTGTGATAAAGCC	*Bam*HI	WSMV cDNA complementary strand 965–982
P1-9-3	TTCC**GAATTC**CCTAGTGCTTGCAAGAATGC	*Eco*RI	WSMV cDNA complementary strand 1442–1461
HPM1-3	CAAC**CTCGAG**TCATTCTAGCATCTTATGTTGC	*Xho*I	WSMV cDNA complementary strand 1663–1681
HPM2-3	ACAC**CTCGAG**TCACTGTCATTTACACTCAGC	*Xho*I	WSMV cDNA complementary strand 1948–1965
HPM3-3	CAAC**CTCGAG**TCAACCCAGATCCTCGCAGAC	*Xho*I	WSMV cDNA complementary strand 2084–2101
HCPRO-3	CCGAATCCTCAGCCAATTTTGTAATCTTT		WSMV cDNA complementary strand 2321–2338
T7-P1C1	*TAATACGACTCACTATAGGGAGACCACATG*TCAAAAACTCAGGAGGAAGAA		WSMV cDNA 638–658
T7-P1C2	*TAATACGACTCACTATAGGGAGACCACATG*GACTTCGATGAAGCTCATGCT		WSMV cDNA 698–718
T7-P1C3	*TAATACGACTCACTATAGGGAGACCACATG*GAATATAAGGAGAAGCAGCTG		WSMV cDNA 755–775
T7-P1C4	*TAATACGACTCACTATAGGGAGACCACATG*CGGATGATGCACTCCAAGGG		WSMV cDNA 821–840
T7-P1C5	*TAATACGACTCACTATAGGGAGACCACATG*GCCAAAATTCCACTGAAGC		WSMV cDNA 884–902
T7-P1C6	*TAATACGACTCACTATAGGGAGACCACATG*GACATACCAGAAGATTTCAGA		WSMV cDNA 944–964

*Italicized letters represent the nucleotide sequences of the promoter for T7 RNA polymerase.

Table 2. PCR templates and primers used for the production of templates for *in vitro* transcription–translation

PCR product*	Template	Forward primer	Reverse primer
A1	pCP1HP-8	T7-ACT	P1-3
A2	pCP1HP-8	T7-ACT	P1-9-3
A3	pCP1HP-8	T7-ACT	HPM1-3
A4	pCP1HP-8	T7-ACT	HPM2-3
A5	pCP1HP-8	T7-ACT	HPM3-3
A6	pCP1HP-8	T7-ACT	HCPRO-3
B1	pCSP1-8	T7-P1C1	HCPRO-3
B2	pCSP1-8	T7-P1C2	HCPRO-3
B3	pCSP1-8	T7-P1C3	HCPRO-3
B4	pCSP1-8	T7-P1C4	HCPRO-3
B5	pCSP1-8	T7-P1C5	HCPRO-3
B6	pCSP1-8	T7-P1C6	HCPRO-3
C1	pCSP1-8	T7-P1C1	HCPRO-3
C2	pCP1-NPT21	T7-P1C1	NPT-3
C3	pCP1-NPT22	T7-P1C1	NPT-3
C4	pCP1-NPT23	T7-P1C1	NPT-3
C5	pCP1-NPT24	T7-P1C1	NPT-3
C6	pCP1-NPT25	T7-P1C1	NPT-3

*DNA templates used for *in vitro* transcription–translation in Fig. 1.

The cDNA fragment from nt positions 914 to 985 in pCP1-F was replaced with nt positions 914 to 2340 of pCSP1-8 to create pCP1HP-8. As the result, pCP1HP-8 contains the 5′-proximal portion of the WSMV polyprotein ORF beginning at the methionine initiation codon (nt 131) to the second nucleotide (nt 2340) of the P3 cistron downstream of the HC-Pro 3′-proximal codon.

A DNA fragment containing the NPT II gene was synthesized by PCR with *P.f.u.* polymerase using pWSMV-JNPT (Choi *et al.*, 2000b) as template and primers X-NPT5 and NPT-3 (Table 1). The NPT II PCR product was digested with *Xho*I and *Sal*I and then inserted into the *Xho*I site of pACT2 to create pACT-XNPT. A nested set of DNA fragments (X21, X22, X23, X24 and X25) from the WSMV P1 cistron was generated by PCR using pACYC-WSMV (Choi *et al.*, 1999) as template with a forward primer (P1-5) and one of the following reverse primers: P1-X21 (for X21), P1-X22 (for X22), P1-X23 (for X23), P1-X24 (for X24) or P1-X25 (for X25); sequences of these and other primers used in this study are given in Table 1. The DNA fragments X21 to X25 were digested with *Xho*I and inserted into the *Xho*I site located between the GAL4 activation domain and NPT II in pACT-XNPT. The resultant plasmids, pCP1-NPT21, -NPT22, -NPT23, -NPT24 and -NPT25, have WSMV P1 sequences fused to NPT II.

Plasmids described above were used as templates for PCR with the primer combinations listed in Table 2 to produce *in vitro* transcription–translation templates. One series of PCR products was based on pCP1HP-8 and shared a common 5′-end with a nested series of 3′-terminal truncations (Fig. 1A). A second series of PCR products was based on pCSP1-8 and shared a common 3′-end with a nested series of 5′-terminal deletions (Fig. 1B). The third series of PCR products represents the P1–NPT II fusions described above (Fig. 1C). *In vitro* transcription–translation was performed in the presence of [^{35}S]methionine using the TNT T7-coupled wheat germ extract system (Promega) as recommended by the manufacturer. DNA template (1 μg) was added to a final volume of 25 μl of transcription–translation mixture and incubated at 30 °C for 2 h. Translation products were analysed by SDS–12% PAGE, followed by autoradiography.

■ **Alignment of potyviral P1 sequences.** Polyprotein amino acid sequences of WSMV and several members of the family *Potyviridae* were aligned using CLUSTAL X. Species of the genus *Potyvirus* compared included TEV (Allison *et al.*, 1986), *Plum pox virus* (PPV) (Maiss *et al.*, 1989) and *Tobacco vein mottling virus* (TVMV) (Domier *et al.*, 1986). Species of the genus *Tritimovirus* included three strains (Sidney 81, Type and El Batán 3) of WSMV (Stenger *et al.*, 1998; Choi *et al.*, 2001) and *Brome streak mosaic virus* (BrSMV) (Götz & Maiss, 1995). Because the P1 cistron is the most variable in length and sequence among members of the family (Ward *et al.*, 1995), the CLUSTAL X alignment required extensive adjustment by eye so that conserved motifs known to be essential for serine proteinase activity (Bazan & Fletterick, 1990; Ryan & Flint, 1997; Verchot *et al.*, 1991, 1992) were brought into alignment.

■ **Insertion of GUS at the P1/HC-Pro junction in an infectious clone of WSMV.** A unique *Sal*I site was engineered in the infectious clone of WSMV (pACYC-WSMV) (Choi *et al.*, 1999) immediately downstream of the codons for the P1 proteinase cleavage site identified in this study (Fig. 2). Three nucleotides (GTC) were inserted between nt positions 1189 and 1190 of pACYC-WSMV by site-directed mutagenesis to produce pWSMV-S1RN. The GUS gene linked to the nucleotide sequence encoding a WSMV NIa proteinase cleavage site was isolated from pWSMV-GUSJ1 (Choi *et al.*, 2000b) and inserted into the *Sal*I site of pWSMV-S1RN to generate pWSMV-GUS-S1RN. *In vitro* transcripts of pWSMV-S1RN and pWSMV-GUS-S1RN were transcribed with SP6 RNA polymerase and inoculated to wheat (*Triticum aestivum* L.) seedlings as described (Choi *et al.*, 1999). Inoculated plants were evaluated for GUS activity by histochemical staining *in situ* (Choi *et al.*, 2000b). Native mobility of GUS produced by pWSMV-GUS-S1RN was assessed following electrophoresis of infected plant protein extracts under non-denaturing conditions by activity gel (zymograph) detection using ELF-97 β-D-glucuronide as described (Steinberg *et al.*, 2000). Purified wild-type GUS protein (Sigma) was used as a positive control in the electrophoretic assay, whereas protein extracted from mock-inoculated or WSMV-S1RN infected plants served as negative controls.

Results and Discussion

Mapping P1 proteolytic domains *in vitro*

The strategy used for mapping domains of the WSMV polyprotein required for P1 proteolysis (Table 2, Fig. 1) was similar to that used for TEV by Verchot *et al.* (1991). A series of 3′-truncations of pCP1HP-8 served as templates for *in vitro* transcription–translation in which the HC-Pro encoding sequence was progressively removed (Fig. 1A). Translation products of the pCP1HP-8 derivatives A1–A6 included a full-length uncleaved polypeptide, the size of which was determined by template length. It was noted that the single major translation product of template A1, truncated at WSMV nt 982 (containing all of the putative P1 domain as proposed by Stenger *et al.*, 1998), was smaller in size than the larger of the two cleaved polypeptides produced by templates A2–A6. The larger of two cleaved polypeptides produced by templates A2–A6 remained constant in size, regardless of additional downstream sequences, and was interpreted as the mature P1

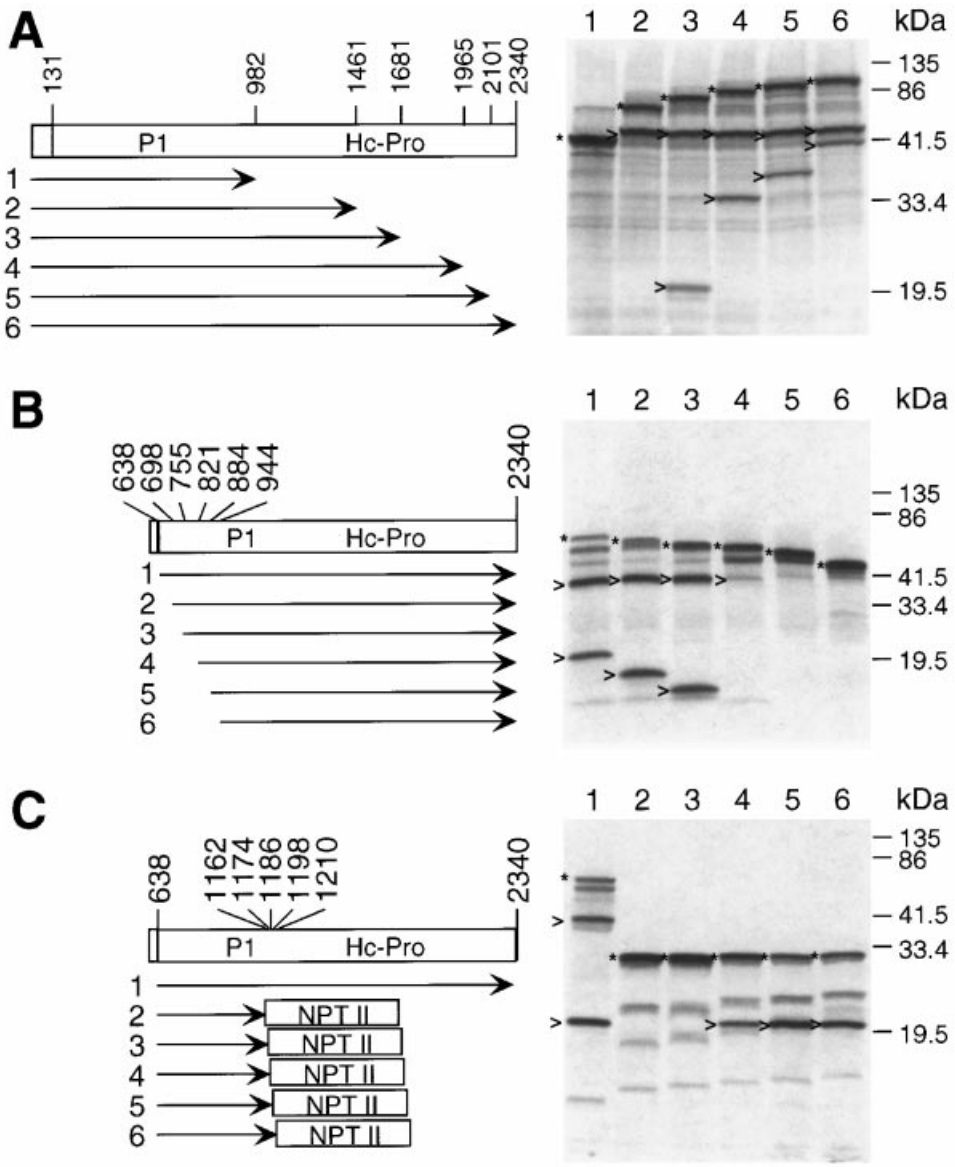

Fig. 1. Mapping of polyprotein domains of *Wheat streak mosaic virus* (WSMV) required for P1 proteolysis. Presented are maps of templates used (at left) and corresponding polypeptides produced (at right) in coupled *in vitro* transcription–translation reactions. Panel (A) depicts a series of 3′-truncations (templates A1–A5) amplified by PCR using pCP1HP-8 (template A6) as template. The region upstream of WSMV nt 131 was derived from pACT2. Panel (B) depicts a series of 5′-deletions (templates B2–B6) amplified by PCR using pCSP1-8 (template B1) as template. Panel (C) depicts a series of templates (C2–C6) in which the 3′-proximal region of pCSP1-8 (template C1) was replaced with NPT II sequences. WSMV nucleotide coordinates are indicated for each map. Mobility and molecular masses of protein standards in kDa are indicated at right. Asterisks (*) denote full-length uncleaved polypeptides, arrowheads (>) denote proteolytically cleaved polypeptides derived from full-length translation products. Unlabelled, fainter bands represent less than full-length polypeptides likely translated by internal initiation at downstream AUG codons.

polypeptide produced by P1 proteolysis. The smaller cleavage product of A3–A6 increased in size, corresponding with template length, and was interpreted as the carboxy-terminal HC-Pro polypeptide released by P1 proteolysis. Although the smaller cleavage product of template A2 was not detected, possibly due to the small size or instability of the polypeptide, detection of the larger cleavage product suggested that P1 proteolysis had occurred. Collectively, these results indicate that P1 proteolysis requires some or all of WSMV nt 983–1461. Therefore, the WSMV P1 domain extends farther downstream than originally proposed (Stenger *et al.*, 1998).

A second series of templates (B1–B6) based on pCSP1-8 bore nested 5′-deletions in the P1 domain (Fig. 1B). For each template, the upstream primer used to amplify the PCR products included an initiation codon to replace the native WSMV polyprotein initiation codon that was removed as part

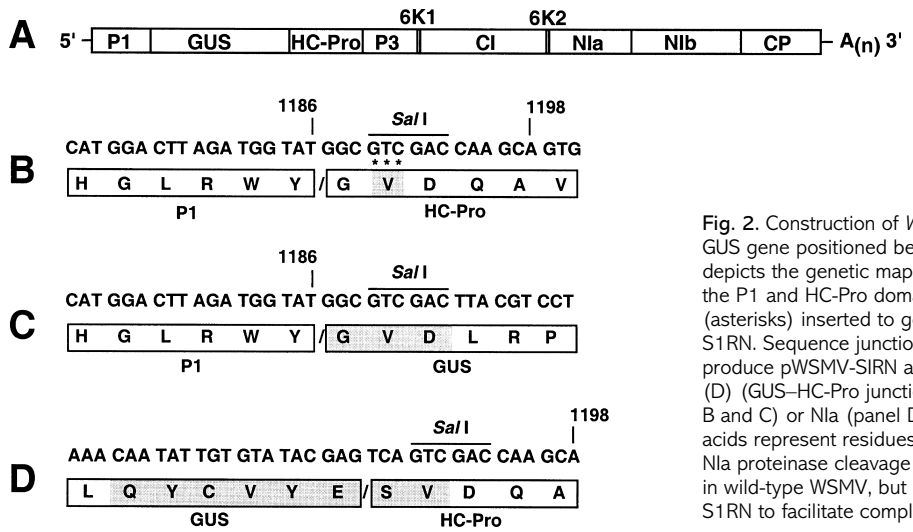

Fig. 2. Construction of *Wheat streak mosaic virus* (WSMV) bearing the GUS gene positioned between the P1 and HC-Pro cistrons. Panel (A) depicts the genetic map of WSMV-Sidney 81 with GUS inserted between the P1 and HC-Pro domains. Panel (B) displays the location of bases (asterisks) inserted to generate a unique *Sal*I restriction site in pWSMV-S1RN. Sequence junctions resulting from insertion of the GUS gene to produce pWSMV-SIRN are depicted in panels (C) (P1–GUS junction) and (D) (GUS–HC-Pro junction). Slashes denote the locations of P1 (panels B and C) or NIa (panel D) proteinase cleavage sites. Shaded amino acids represent residues not present in wild-type proteins. Note that the NIa proteinase cleavage site (QYCVYE/S) is not present in this location in wild-type WSMV, but was engineered at this position in pWSMV-GUS-S1RN to facilitate complete excision of GUS from the polyprotein.

of the 5′-deletion. In this experiment, the uncleaved full-length translation product became progressively smaller in size, corresponding to the size of the deletion. For templates B1–B3, two prominent cleaved polypeptide products were observed. The larger cleaved product remained constant in size, and was interpreted as being the carboxy-terminal HC-Pro polypeptide. The smaller cleavage product varied in size according to the extent of the 5′-deletion and was interpreted as the cleaved P1 polypeptide. The translation product of template B4 had impaired proteolytic activity, with only minor amounts of the large cleavage product detected. No evidence of proteolysis was detected for the translation products of templates B5 and B6. These results indicated that the proteolytic domain and cleavage site of P1 reside downstream of nt 755 (amino acid residue 208). Thus, as is the case for TEV, the proteolytic domain is located in the carboxy-terminal region of P1.

To fine map the P1 proteinase cleavage site, a third series of *in vitro* transcription–translation templates (C2–C6) bearing WSMV sequences from nt 638 to 1162–1210 was constructed in which downstream HC-Pro sequences were replaced with NPT II sequences (Fig. 1C). Translation of template C1 (retaining WSMV sequences from nt 638 through the 3′-proximal end of HC-Pro) produced two cleavage products, the larger of which corresponded to full-length HC-Pro and a smaller product represented the carboxy-terminal portion of P1. Apparently, templates C2 and C3 did not encode the intact proteinase cleavage site and produced only uncleaved translation products. In contrast, templates C4–C6 encoded the functional proteinase cleavage site, thereby producing cleaved polypeptides corresponding to the carboxy-terminal portion of P1. Unexpectedly, the second predicted polypeptide cleavage product corresponding to NPT II sequences was not detected. Furthermore, sizes of the uncleaved translation products were less than expected for the P1–NPT II fusions. Sequence determined for the NPTII insert of pACT-XNPT revealed an unintended frame-shift mutation in the NPT II sequence, such that the P1–NPT II translation products would be prematurely terminated to yield a full-length uncleaved polypeptide of ~29 kDa. The small size (9 kDa) of the predicted carboxy-terminal polypeptide released by P1 proteolysis explains why the second proteolytic cleavage product was not detected by autoradiography. Nonetheless, these results indicate that HC-Pro sequences are not required for P1 proteolysis and that the P1 cleavage site of WSMV occurs downstream of residue G_{348} and upstream of residue G_{353}.

Hallmark serine proteinase motifs are present in WSMV P1

Collectively, the *in vitro* transcription–translation experiments indicated that the WSMV P1 proteinase cleavage site resides considerably downstream from the $YS_{283-284}$ dipeptide. CLUSTAL X failed to align the known P1 proteinase cleavage site of TEV with the region of the WSMV polyprotein implicated by the *in vitro* transcription–translation experiments. This is not surprising considering that the P1 domain is the least conserved and most variable in length among monopartite viruses of the family *Potyviridae* (Ward *et al.*, 1995) However, when the WSMV Y_{352} residue was manually aligned with the Y_{304} residue of TEV, upstream amino acid residues required for TEV proteolysis came into alignment with similar residues present in WSMV (Fig. 3). Specifically, the WSMV polyprotein contained H_{257}, D_{267} and S_{303} that were conserved at similar positions in the modified alignment, relative to all other taxa examined. These three residues are required for proteinase activity of TEV P1 (Verchot *et al.*, 1991, 1992) and likely constitute the catalytic H-D-S triad associated with serine proteinases (Bazan & Fletterick, 1990). The serine residue was flanked by glycine residues (GXSG), another hallmark of the serine proteinase H-D-S catalytic triad (Ryan & Flint, 1997). However, the spacing between the H, D and S residues is somewhat more compact for potyviruses and tritimoviruses

```
                  2         2               22                  3
                  5         6               88                  0
                  7         7               34                  3
                  *         *                                   *
TEV               HMYGERKR-VDLRIDNWQQETLLDLAKRFKNERVDQ---SKLTFGSSGLVLR-QGSYGPA
PPV               HLDGSKPR-YDLVLDEATKKILQLFANTSGFHHVHKKG--EVTPGMSGFVVNPMNLSDPM
TYMV              HAKGHRRR-IDCRMHRREQRTMHMFMRKTTKTEVRSK---HLRKGDSGIVLLTQKIKGHL
WSMV-Sidney 81    HTMGYPKRDWDATKDIPEDFRGFITTYSGVIQYTRKVQDHEVTLGWSGVLLSEMDVPDGY
WSMV-Type         HTMGYPKRDWDASKDIPEDLRGFITTYSGVIQYTRKVQDHEVTLGWSGVLLSEMDVPDGY
WSMV-El Batán 3   HTMGYPKREWDGCKDVPEDMRSFITTYSSVIQYTRKVQDHEITLGWSGVVLSEFDVPDGY
BrSMV             H--VLESNNVDPSDDLYADVVPFLKHYGTCGRPVGVINPRDIKPGWSGVVLLQDELPESL

                          3   3                       3 3
                          2   2                       5 5
                          5   9                       2 3
                          ^^^^^                        /
TEV               H-WYRH-GMFIVRGRSDGMLVDARAKVTFAVCHSMTHY/SD  (304 aa)
PPV               -QVYDTD-LFIVRGKHNSILVDSRCKVSKKQSNEIIHY/SD  (308 aa)
TVMV              SGV-R-DEFFIVRGTCDDSLLEARARFSQSITLRATHF/ST  (274 aa)
WSMV-Sidney 81    -QEDCVDGLFIVMGRCAHGRIQNALKP--KCTHGLRWY/GD  (352 aa)
WSMV-Type         -QEDCVDGLFIVMGRCAHGRIQNALKP--KCTHGFRWY/GD  (352 aa)
WSMV-El Batán 3   -QEDCVDGLFIVMGRCAHGRIQNALKP--RCIHGLRWY/GD  (352 aa)
BrSMV             HQE-CVDGVFVVQGIGPDGQLKNALKT--TTGERIEYY/SS  (403 aa)
```

Fig. 3. Alignment of potyvirus P1 carboxy-terminal domains. Presented are partial polyprotein sequences of three species of the genus *Potyvirus* [*Tobacco etch virus* (TEV), *Tobacco vein mottling virus* (TVMV) and *Plum pox virus* (PPV)], and two species of the genus *Tritimovirus* [*Wheat streak mosaic virus* (WSMV) strains Sidney 81, Type and El Batán 3 and *Brome streak mosaic virus* (BrSMV)]. Amino acid coordinates indicated are for WSMV. Asterisks (*) denote conserved histidine (H), aspartic acid (D) and serine (S) residues required for TEV P1 proteolysis. Hat symbols (^) denote location of motifs similar or identical to the FIVRG motif also required for TEV P1 proteolysis. Slash (/) denotes location of the P1 peptide bond hydrolysis delimiting the carboxy-terminal domain of P1 from the amino-terminal domain of HC-Pro. Numbers in parentheses indicate the number of amino acid residues in the mature P1 protein of each virus.

than that of the serine proteinase encoded by ORF 1 of the polerovirus *Potato leafroll virus* (Sadowy *et al.*, 2001) or other serine proteinases of the picornavirus super-group (Ryan & Flint, 1997). Additionally, the WSMV motif FIVMG$_{325–329}$ aligned with a similar motif in TEV, TVMV and PPV (FIVRG), and BrSMV (FVVQG), that is also required for TEV P1 proteolysis (Verchot *et al.*, 1991). Other than these motifs, there was little conservation of intervening sequences among different species of the family *Potyviridae* (Fig. 3).

The templates B1–B3 bearing deletions in the 5′-proximal region of WSMV P1 upstream of the H-D-S triad retained proteinase activity, whereas extension of the deletion to include loss of both H_{257} and D_{267} (template B6) resulted in no detectable proteinase activity (Fig. 1B). However, because templates B4 and B5 contained the H-D-S triad but exhibited reduced (B4) or no (B5) proteinase activity, sequences upstream (nt 755–884) of the H-D-S triad are also required. This observation is consistent with mutational analysis of the TVMV P1 proteinase, in which insertion of four codons in the same region upstream of the H-D-S triad (between TVMV amino acid residues 155 and 156) abolished proteolytic activity (Moreno *et al.*, 1999).

The WSMV P1 domain includes all known essential motifs of a serine proteinase with the probable site of peptide bond hydrolysis between Y_{352} and G_{353}, rather than between Y and S residues as in TEV. Thus, mature WSMV P1 protein would be 48 amino acid residues longer than P1 of TEV. The P1 proteinase cleavage site predicted for a second tritimovirus, BrSMV (Götz & Maiss, 1995; Fig. 3), would yield a longer P1 polypeptide (403 amino acids) with hydrolysis between Y and S residues. However, differences between species of the same genus may not be unusual, as the potyvirus TVMV has S_{275} as the amino-terminal residue of HC-Pro, indicating that P1 cleavage occurs between F and S residues (Mavankal & Rhoads, 1991). Thus, although monopartite species of the family *Potyviridae* retain serine proteinase activity to cleave the N-terminal region of the polyprotein, the length of the mature P1 polypeptide varies, as does the residues between which the peptide bond is hydrolysed.

WSMV P1 proteinase functions *in vivo*

Dolja *et al.* (1992) inserted GUS between the P1 and HC-Pro domains of TEV without destroying infectivity so long as the polyprotein was proteolytically processed to remove P1. Although insertion of a 12 base sequence containing a *Sal*I site between WSMV nt 991 and 992 did not affect infectivity, subsequent insertion of GUS at the engineered S*al*I site in WSMV abolished infectivity (Choi *et al.*, 2000b). In light of the present study, these results are not surprising, as elements of the catalytic domain were separated by over 600 amino acid residues due to insertion of GUS. However, since the initial 12 base insertion retained infectivity, evidently some variation in the spacing of critical motifs required for P1 proteinase is allowed.

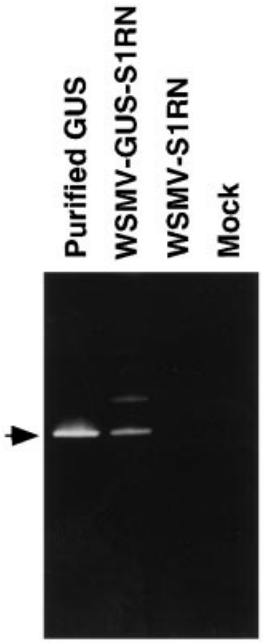

Fig. 4. Detection of active GUS protein. Protein samples electrophoresed in polyacrylamide gels under non-denaturing conditions were incubated with GUS substrate and detected by UV fluorescence. Protein samples were wild-type GUS protein obtained from a commercial source ('Purified GUS'), or total soluble protein extracted from mock-inoculated wheat ('Mock'), or wheat infected with WSMV lacking ('WSMV-S1RN') or bearing ('WSMV-GUS-S1RN') GUS. Arrow at left denotes mobility of purified wild-type GUS protein.

To provide additional evidence that the P1 proteinase cleavage site of WSMV occurs between amino acid residues Y_{352} and G_{353}, the GUS reporter gene was repositioned adjacent to and downstream of this dipeptide in the infectious clone of WSMV. Excision of GUS from the viral polyprotein would occur by P1 proteolysis at the amino end and by NIa-catalysed cleavage at the carboxy end at an engineered NIa proteinase cleavage site (Fig. 2). *In vitro* transcripts derived from the GUS-containing construct pWSMV-GUS-S1RN produced systemic infection of wheat and GUS activity was demonstrated in systemically infected wheat leaves by a histochemical assay (data not shown). Active GUS protein also was detected in a total protein sample extracted from WSMV-GUS-S1RN-infected plants and subjected to PAGE under non-denaturing conditions (Fig. 4). A substantial fraction of active GUS protein extracted from pWSMV-GUS-S1RN-infected plants exhibited electrophoretic mobility similar to that observed for purified wild-type GUS protein obtained from a commercial source. This result confirms that GUS protein was excised from the polyprotein and provides additional *in vivo* evidence that the WSMV P1 proteinase cleavage site was correctly identified. The presence of some active GUS protein in a second, slower migrating band unique to pWSMV-GUS-S1RN indicated that proteolysis was incomplete. Incomplete proteolysis may have resulted from the co-translational requirement for P1 autoproteolysis (Verchot *et al.*, 1992), and/or alteration of local topology of the cleavage site in which native HC-Pro sequences were replaced with GUS sequences. Nonetheless, identification of the P1 proteinase cleavage site permitted insertion of a foreign gene sequence at a second genomic location in WSMV without compromising infectivity and may expand the utility of WSMV as a gene expression vector in monocotyledonous hosts.

We thank David Dunigan, Les Lane and Dallas L. Seifers for helpful comments and suggestions. Mention of proprietary or brand names is necessary to report factually on available data; however, the USDA neither guarantees nor warrants the standard of the product, and the use of the name by USDA implies no approval to the exclusion of others that also may be suitable. This article is in the public domain and not copyrightable. It may be freely reprinted with customary crediting of source.

References

Allison, R. F., Johnson, R. E. & Dougherty, W. G. (1986). The nucleotide sequence of the coding region of tobacco etch virus genomic RNA: evidence for the synthesis of a single polyprotein. *Virology* **154**, 9–20.

Bazan, J. F. & Fletterick, R. J. (1990). Structural and catalytic models of trypsin-like proteases. *Seminars in Virology* **1**, 311–322.

Carrington, J. C. & Dougherty, W. G. (1987). Small nuclear inclusion protein encoded by a plant potyvirus genome is a protease. *Journal of Virology* **61**, 2540–2548.

Carrington, J. C., Cary, S. M., Parks, T. D. & Dougherty, W. G. (1989). A second proteinase encoded by a plant potyviral genome. *EMBO Journal* **8**, 365–370.

Choi, I.-R., French, R., Hein, G. L. & Stenger, D. C. (1999). Fully biologically active *in vitro* transcripts of the eriophyid mite-transmitted wheat streak mosaic tritimovirus. *Phytopathology* **89**, 1182–1185.

Choi, I.-R., Stenger, D. C. & French, R. (2000a). Multiple interactions among proteins encoded by the mite-transmitted wheat streak mosaic tritimovirus. *Virology* **267**, 185–198.

Choi, I.-R., Stenger, D. C., Morris, T. J. & French, R. (2000b). A plant virus vector for systemic expression of foreign genes in cereals. *Plant Journal* **23**, 547–555.

Choi, I.-R., Hall, J. S., Henry, M., Zhang, L., Hein, G. L., French, R. & Stenger, D. C. (2001). Contributions of genetic drift and negative selection on the evolution of three strains of wheat streak mosaic tritimovirus. *Archives of Virology* **146**, 619–628.

Dolja, V. V., McBride, H. J. & Carrington, J. C. (1992). Tagging of plant potyvirus replication and movement by insertion of β-glucuronidase into the viral polyprotein. *Proceedings of the National Academy of Sciences, USA* **89**, 10208–10212.

Domier, L. L., Franklin, K. M., Shahabuddin, M., Hellmann, G. M., Overmeyer, J. H., Hiremath, S. T., Siaw, M. F., Lomonossoff, G. P., Shaw, J. G. & Rhoads, R. E. (1986). The nucleotide sequence of tobacco vein mottling virus RNA. *Nucleic Acids Research* **14**, 5417–5430.

Götz, R. & Maiss, E. (1995). The complete nucleotide sequence and genome organization of the mite-transmitted brome streak mosaic rymovirus in comparison to those of potyviruses. *Journal of General Virology* **76**, 2035–2042.

Maiss, E., Timpe, U., Brisske, A., Jelkmann, W., Casper, R., Himmler, G., Mattanovich, D. & Katinger, H. W. (1989). The complete nucleotide sequence of plum pox virus RNA. *Journal of General Virology* **70**, 513–524.

Mavankal, G. & Rhoads, R. E. (1991). In vitro cleavage at or near the N-terminus of the helper component protein in tobacco vein mottling virus polyprotein. *Virology* **185**, 6110–6114.

Moreno, M., Brandwagt, B. F., Shaw, J. G. & Rodríguez-Cerezo, E. (1999). Infectious virus in transgenic plants inoculated with a nonviable P1-proteinase defective mutant of a potyvirus. *Virology* **257**, 322–329.

Ryan, M. D. & Flint, M. (1997). Virus-encoded proteinases of the picornavirus super-group. *Journal of General Virology* **78**, 699–723.

Sadowy, E., Juszczuk, M., David, C., Gronenborn, B. & Hulanicka, M. D. (2001). Mutational analysis of the proteinase function of *Potato leafroll virus*. *Journal of General Virology* **82**, 1517–1527.

Steinberg, T. H., Lauber, W. M., Berggren, K., Kemper, C., Yue, S. & Patton, W. F. (2000). Fluorescence detection of proteins in sodium dodecyl sulfate–polyacrylamide gels using environmentally benign, nonfixative saline solution. *Electrophoresis* **21**, 497–508.

Stenger, D. C., Hall, J. S., Choi, I.-R. & French, R. (1998). Phylogenetic relationships within the family *Potyviridae*: wheat streak mosaic virus and brome streak mosaic virus are not members of the genus *Rymovirus*. *Phytopathology* **88**, 782–787.

Verchot, J., Koonin, E. V. & Carrington, J. C. (1991). The 35-kDa protein from the N-terminus of the potyviral polyprotein functions as a third virus-encoded proteinase. *Virology* **185**, 527–535.

Verchot, J., Herndon, K. L. & Carrington, J. C. (1992). Mutational analysis of the tobacco etch potyviral 35-kDa proteinase: identification of essential residues and requirements for autoproteolysis. *Virology* **190**, 298–306.

Ward, C. W., Weiller, G. F., Shukla, D. D. & Gibbs, A. (1995). Molecular systematics of the *Potyviridae*, the largest plant virus family. In *Molecular Basis of Virus Evolution*, pp. 477–500. Edited by A. Gibbs, C. Calisher & F. García-Arenal. Cambridge: Cambridge University Press.

Received 23 August 2001; Accepted 26 October 2001